Self-supervised nonlocal spectral similarity-induced material decomposition network for dual-energy CT

Lei Wang[a,b], Yongbo Wang[a,b], Zhaoying Bian[a,b], Dong Zeng[a,b], and Jianhua Ma[a,b]

[a]Department of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[b]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China

ABSTRACT

Dual-energy computed tomography (DECT) imaging plays an important role in clinical diagnosis applications due to its material decomposition capability. However, in the cases of low-dose DECT imaging and ill-conditioned issue, the direct decomposed material images from DECT images would suffer from severe noise-induced artifacts, leading to low quality and accuracy. In this paper, we propose a self-supervised Nonlocal Spectral Similarity-induced Decomposition Network (NSSD-Net) to produce decomposed material images with high quality and accuracy in the low-dose DECT imaging. Specifically, we first build the model-driven iterative decomposition model and optimize the objective function by the iterative shrinkage-thresholding algorithm (ISTA) with the convolutional neural network. Considering the intrinsic characteristics information (i.e., structural similarity and spectral correlation) underlying DECT images, which can be used as the prior information to improve the accuracy of the decomposed material images, we construct the nonlocal spectral similarity-based cost function by using the prior information and incorporating it into the iterative decomposition network to guarantee stability. The proposed NSSD-Net method was validated and evaluated with real clinical data. Experimental results showed that the presented NSSD-Net method outperforms the other competing methods in terms of noise-induced artifacts reduction and decomposition accuracy.

Keywords: Dual-energy computed tomography, deep learning, material decomposition, low-dose, self-supervised learning

1. INTRODUCTION

Dual-energy computed tomography (DECT) has been widely used in clinical diagnosis, including kidney stone characterization, iodine quantitative examination, and many other applications. Compared to traditional CT, DECT provides two sets of attenuation measurements by exploiting two different energy spectra to achieve the material decomposition and energy-selective imaging, which opens up new diagnosis possibilities. However, the direct decomposed material images from DECT images would suffer from severe noise and noise-induced artifacts, especially under low-dose scan protocols. The main reason is that the material decomposition process suffers an ill-conditioned problem that leads to severe noise in the decomposed material images.[1] To tackle this problem, many advanced algorithms have been proposed to improve the signal-to-noise ratio of decomposed material images.

These material decomposition algorithms can be generally divided into two categories: model-driven material decomposition methods and data-driven material decomposition methods. The model-driven methods[2] generally build a model on the physical properties and the attenuation characteristics, then optimize the objective function to get the decomposed material images. These methods also can be characterized into one-step inversion, projection-domain, and image-domain decomposition methods. One-step inversion is mathematically the most elegant with one-step matrix inversion, but they are very computationally expensive. Projection-domain methods directly decompose the projection data into the basis material sinograms and then reconstruct them with the filter back-projection algorithm. Although the decomposition performance is significant, they require accurate

Correspondence: Jianhua Ma.: E-mail: jhma@smu.edu.cn

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 123040W
 · doi: 10.1117/12.2646645

system calibrations that use nonlinear models. Image-domain methods reconstruct each energy bin image and then directly perform material decomposition on image data, which may lead to beam hardening artifacts. On the other hand, the data-driven methods[3] train the convolutional neural network (CNN) in an end-to-end manner to obtain accurate decomposed material images. However, these methods have two main problems. First, these methods do not take into account the inherent physics mechanism, their unexplained property obstacles deep learning techniques to be widely applied in clinical. Second, most of these existing algorithms are supervised learning methods. The decomposition performance depends on the quality and quantity of training data pairs. However, high-quality training data pairs are difficult to acquire in clinic, which limits the accuracy of decomposed material images and the generalization performance of models.

To address these problems, we propose a nonlocal spectral similarity-induced material decomposition network (NSSD-Net) for DECT. The NSSD-Net couples the model-based material decomposition model with the deep network and optimizes by the iterative shrinkage-thresholding algorithm (ISTA). Considering the structural similarity and spectral correlation underlying DECT images, the nonlocal spectral similarity-based cost function is designed to guarantee network convergence in a self-supervised manner. The main contributions of the proposed NSSD-Net can be summarized as follows: the first one is that the proposed method fully considers the material decomposition physician mechanism and combines the benefits of model-driven methods and data-driven methods. The second one is that the proposed methods use the nonlocal spectral similarity features as prior information and optimize hyper-parameters by self-supervised learning strategy with no ground-truth data, which enhances the generalization of the proposed model. Experiments on clinical data have confirmed the significant decomposition performance of the NSSD-Net.

2. METHODS

2.1 Model-driven material decomposition model

According to image-based decomposition theory, the linear combination of pixel values in the basis material images can represent the linear attenuation coefficient of each pixel in the input images. Based on the above assumption, the formulation of the material composition for DECT can be written as:

$$U = AX, \tag{1}$$

where $U = [\mu_H, \mu_L]^T$ denotes the linear attenuation coefficient of high and low energy spectrum, $X = [x_1, x_2]^T$ denotes the normalized densities of the basis materials, A is the decomposition matrix and can be represented as follows:

$$A = \begin{pmatrix} \mu_{M1H} I & \mu_{M2H} I \\ \mu_{M1L} I & \mu_{M2L} I \end{pmatrix}, \tag{2}$$

where μ_{Mij} denotes the mass attenuation coefficient of the basis materials, I is an identity matrix with the dimension of $2N$-by-$2N$, and N is the total number of pixels in one CT image.

Direct decomposition generating decomposed material images with severely degraded SNR, especially at low-dose levels. This is because there is a large condition number on the matrix A, and this can be treated as an ill-posed problem, which makes the decomposition process is sensitive to the noise on the raw CT images. To alleviate this situation, least-square estimation model with regularization $\mathcal{R}(X)$ is introduced to suppress decomposed material images noise, which can be expressed as follows:

$$X^* = \underset{X}{\operatorname{argmin}} \|AX - U\|_2^2 + \lambda \mathcal{R}(X), \tag{3}$$

where $\mathcal{R}(X) = \|X\|_1$ and λ is the hyper-parameters of regularization term.

ISTA is a prevailing framework to optimize the decomposition solvation process with non-smooth regularizers. Each iteration of ISTA involves gradient descent update the decomposed material images followed by a shrinkage-threshold step:

$$X^{(k)} = \arg\min_{X} \left\| X - R^{(k)} \right\|_2^2 + \lambda \|X\|_1, \tag{4}$$

$$R^{(k)} = X^{(k-1)} - \rho A^T \left(AX^{(k-1)} - U \right), \tag{5}$$

where ρ is the step size, $X^{(k)}$ and $R^{(k)}$ denote the intermediate variables of the decomposed results, and λ is the hyper-parameter of the regularization term. However, the model-based material decomposition methods are computationally expensive, and it is difficult to select the optimal parameters (i.e., ρ and λ) and the global optimal solution in practical applications.

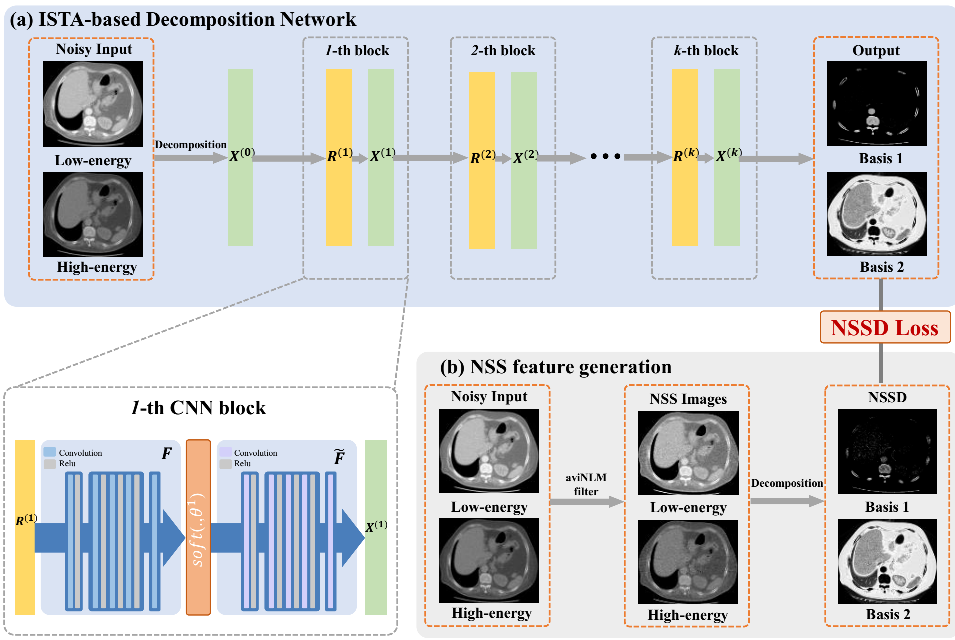

Figure 1. The overall framework of the proposed NSSD-Net.

2.2 The proposed NSSD-Net

To address these problems, we propose a self-supervised nonlocal spectral similarity-induced material decomposition network. As shown in Fig. 1, the proposed NSSD-Net consists of two parts: (a) ISTA-based decomposition network and (b) NSS feature generation.

(a) ISTA-based decomposition network: Inspired by the powerful representational capability and universal approximation of CNN, we use the CNN model $F(\cdot)$ stands for the material decomposition nonlinear function to suppress the noise of images. Thus, the ISTA-based material decomposition model can be formulated as:

$$X^{(k)} = \arg\min_{X} \left\| F(X) - F\left(R^{(k)}\right) \right\|_2^2 + \theta \|F(X)\|_1, \tag{6}$$

$$R^{(k)} = X^{(k-1)} - \rho A^T \left(AX^{(k-1)} - U \right), \tag{7}$$

where θ is a learnable parameter. Therefore, the k-th iteration of the decomposed material images $X^{(k)}$ can be efficiently computed in closed-form as follows:

$$X^{(k)} = \tilde{F}\left(\text{soft}\left(F\left(R^{(k)}\right), \theta^{(k)}\right)\right), \tag{8}$$

where $\tilde{F}(\cdot)$ denotes the left inverse of $F(\cdot)$, and has a symmetrical structure with $F(\cdot)$.

(b) NSS feature generation: Considering the intrinsic characteristics underlying DECT images, i.e., structural similarity and spectrum correlation, which characterizes the priors of desired DECT images. Based on the success of the previous work,[4] in this work, we utilize this prior information to generate the nonlocal spectral similarity (NSS) features to enhance the stability of the ISTA-based iterative decomposition network. Specifically, the proposed method is designed by averaging the acquired DECT images $U = [\mu_H, \mu_L]^T$ to obtain the average image U_{avi}, which takes the image similarity within the two energies into consideration. Then, we use the aviNLM method[4] to denoise the noisy DECT images U and estimate the decomposed material image X_p by the image-domain-based decomposition algorithm. The decomposed material image X_p can provide the prior information to guide the ISTA-based iterative decomposition network. Thus, the proposed nonlocal spectral similarity-based cost function can be represented as:

$$L_{NSSD} = \frac{1}{N} \sum_{i=1}^{N} \left\| {X_i}^{(N_t)} - X_p \right\|_2^2, \tag{9}$$

where N and N_t are the number of training data and basic CNN module, respectively. ${X_i}^{(N_t)}$ is the decomposed result of the NSSD-Net.

2.3 Loss Function

NSSD-Net takes the noisy DECT image $\{U_i\}_{i=1}^{N}$ and NSS feature-induced decomposed material image $\{X_p\}_{i=1}^{N}$ as input, and generates the decomposed material image, denoted by ${X_i}^{(N_t)}$ as output. We optimize the proposed network by the self-supervised learning strategy and the total loss function can be expressed as:

$$L_c = \frac{1}{N_t N} \sum_{i=1}^{N} \sum_{k=1}^{N_t} \left\| \tilde{F}^{(k)} \left(F^{(k)} \left(X_i \right) \right) - X_p \right\|_2^2, \tag{10}$$

$$L_{\text{total}} = L_{NSSD} + \tau L_c. \tag{11}$$

where τ is a constant and we set $\tau = 1$ in this work.

2.4 Experimental Dataset and Settings

The clinical patient study was conducted to assess the proposed NSSD-Net performance. After writing the informed consent from 29 volunteer patients, clinical data were acquired from the GE Discovery CT750 HD scanner with 140kVp and 80kVp. The virtual monochromatic spectral (VMS) images were generated using the GE commercial software. In this work, the VMS images at 90kev and 140kev were selected as the DECT images. In addition, the linear attenuation coefficient of basis materials according to the VMS images reported by the National Institute of Standards and Technology (NIST). Then, we simulated the noisy DECT data by the previous work[5] to further analyze the noise suppression performance among different material decomposition methods.

In the experiment, 1000 simulated noisy VMS image pairs (i.e., 90kev and 140kev) from 22 patients were used for training and 188 simulated noisy VMS images pairs from the remaining patients were used for testing. The network was implemented on an NVIDIA Tesla P40 GPU based on the PyTorch framework and using the Adam optimizer. The initial learning rate, batch size, epoch, and iterations were set to $1e^{-3}$, 1, 50, and 100, respectively. During training, the number of CNN basic module was set to 8. According to experience, the initial gradient descent step $\rho = 0.5$, the hyper-parameter $\tau = 1$, and the threshold $\theta = 0.1$.

3. RESULTS

Fig. 2 shows the decomposed material images of different methods for two test cases. The first column is the ground-truth decomposed material images. It can be observed that both direct matrix inversion method (DIMD) and statistic iterative based decomposition method (Iterative MD) failed to properly distinguish basis materials, and the decomposed material images were heavily corrupted by the noise and noise-induced artifacts. Compared to the DIMD and Iterative MD, the data-based material decomposition methods (i.e., CD-Convnet and ISTA-Net) remove the noise and artifacts further and provide accurately estimates of different materials. However, compared to NSSD-Net, the data-based material decomposition methods tend to over-smoothen the images,

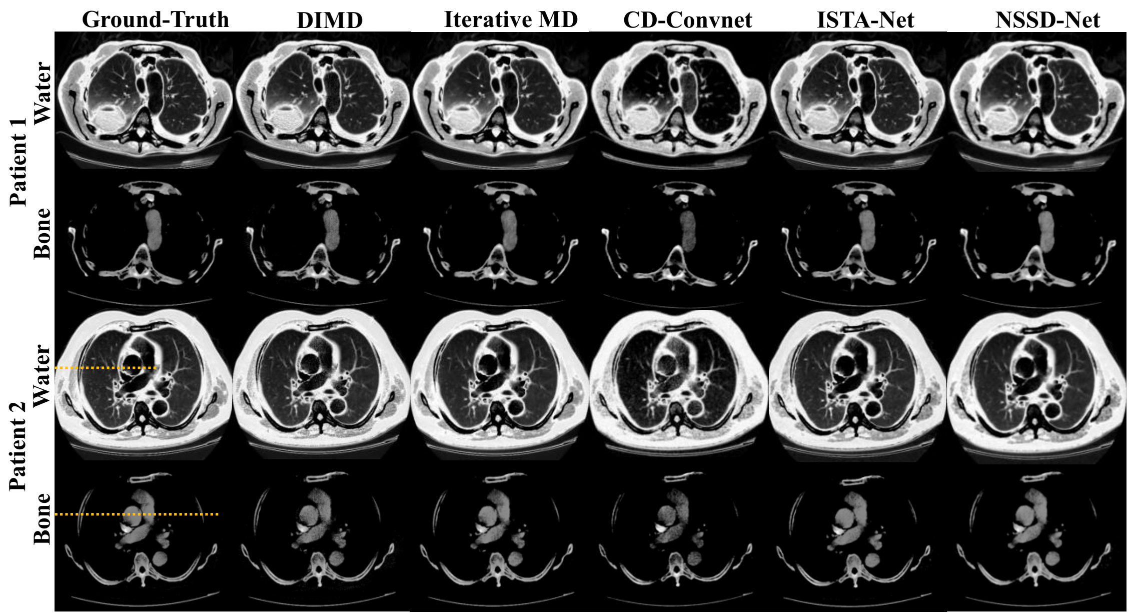

Figure 2. The decomposed material images of different methods: the first and third row shows the water decomposed images with the display window [0, 1] $g.cm^{-3}$; the second and fourth row shows the bone decomposed images with the display window [0.5 1.5] $g.cm^{-3}$.

especially in the structural edge and soft tissue area. The results demonstrate that the self-supervised nonlocal spectral similarity prior can provide the spectral data structure feature information and effectively guide the material decomposition process. Fig. 3 shows the profile comparison indicated by the orange lines in Fig. 2. It can be judged that the NSSD-Net achieves a good balance between bias control and noise suppression.

In order to quantitatively validate the decomposed performance of the proposed NSSD-Net, the peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM) and root mean square error (RMSE) metrics are calculated. As shown in Tab. 1, both NSSD-Net results (i.e. water and bone equivalent fractions), achieve relative superior quantitative metrics than the other decomposition methods expect for the ISTA-Net. The possible reason is that the ISTA-Net utilize the fully supervised manner to train the network, while the proposed NSSD-Net trained without the labeled data. It is worth noting that the proposed method outperforms the ISTA-Net in terms of the SSIM and confirm that the NSSD-Net can provide the global and local structure information from DECT images.

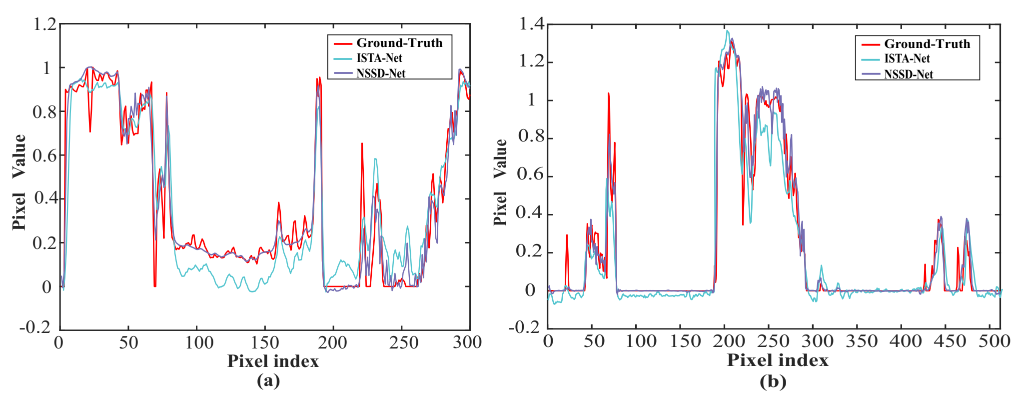

Figure 3. Profile comparison of the ISTA-Net and NSSD-Net along the orange line indicated in the Fig. 2: (a) water decomposed image, (b) bone decomposed image.

Table 1. Quantitative results of different material decomposition methods.

Methods		PSNR	SSIM	RMSE
DIMD	bone	22.1648 ± 1.8508	0.7946 ± 0.0995	8.8992 ± 3.0836
	water	21.6688 ± 2.0852	0.8072 ± 0.0866	9.1729 ± 2.9362
Iterative MD	bone	27.6527 ± 1.0160	0.9103 ± 0.0392	3.6492 ± 1.1476
	water	24.0065 ± 0.8603	0.8932 ± 0.0385	4.3312 ± 1.0572
CD-Convnet	bone	28.1554 ± 0.8808	0.9264 ± 0.0198	4.0299 ± 0.8422
	water	29.2132 ± 0.3138	0.8647 ± 0.0365	6.1902 ± 2.0852
ISTA-Net	bone	$\mathbf{32.5106 \pm 1.0177}$	0.9602 ± 0.0142	$\mathbf{1.4208 \pm 0.6827}$
	water	$\mathbf{32.6545 \pm 0.7508}$	0.9684 ± 0.0251	$\mathbf{2.6060 \pm 0.8472}$
NSSD-Net	bone	31.4440 ± 1.0940	$\mathbf{0.9631 \pm 0.0215}$	1.6329 ± 0.6989
	water	31.3831 ± 0.8861	$\mathbf{0.9701 \pm 0.0127}$	2.9008 ± 0.5268

4. CONCLUSION

In this paper, we propose a novel material decomposition framework for DECT based on the self-supervised nonlocal spectral similarity prior. Specifically, the proposed NSSD-Net combines the traditional model-based decomposition model with the data-based method and optimizes the objective function by the nonlocal spectral similarity prior. The experimental results with clinical data show that the NSSD-Net can obtain accurate decomposed material images of dual-energy.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant U21A6005 and Grant U1708261, the National Key R&D Program of China under Grant No.2020YFA0712200, and Young Talent Support Project of Guangzhou Association for Science and Technology.

REFERENCES

[1] W. A. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual energy CT material density images." *IEEE Transactions on Medical Imaging*, vol. 7, no. 3, pp. 218-224, 1988.

[2] N. Ducros, J. Abascal, and B. Sixou, "Regularization of nonlinear decomposition of spectral x-ray projection images." *Medical Physics*, vol. 44, no. 9, pp. e174-e187, 2017.

[3] Z. Shi, H. Li, and Q. Cao, "A material decomposition method for dual-energy CT via dual interactive Wasserstein generative adversarial networks." *Medical Physics*, vol. 48, no. 6, pp. 2891-2905, 2021.

[4] H. Zhang, D. Zeng, J. Lin, and J. Ma, "Iterative reconstruction for dual energy CT with an average image-induced nonlocal means regularization." *Physics in Medicine and Biology*, vol. 62, no. 13, pp. 5556-5574, 2017.

[5] D. Zeng, J. Huang, Z. Bian, and J. Ma, "A Simple Low-Dose X-Ray CT Simulation From High-Dose Scan." *IEEE Transactions on Nuclear Science*, vol. 62, no. 5, pp. 2226-2233, 2015.